

文章编号: 2095-4980(2020)01-0150-05

基于图结构的城市出租车交通流量可视化

张伟明¹, 张勇^{1,2}, 刘浩², 孙艳丰¹

(1.北京工业大学 信息学部, 北京 100124; 2.北京市交通信息中心, 北京 100161)

摘要: 为了更好地分析出租车在城市区域中的运行规律, 提出一种基于图结构的城市交通流量可视化分析方法。通过对路网进行聚类, 将路网连接路段划分为区域结构, 并用点-线连接形式表示城市路网, 同时以区域车流量为权重, 结合图中心性概念对区域重要性进行了分析。以北京市四环为例, 对一亿多条出租车 GPS 数据进行可视化分析, 实验结果表明, 该方法可以直观有效地展现不同区域的出租车流量随时间变化规律和不同区域的重要程度。

关键词: 出租车 GPS; 图结构; 流量; 可视化

中图分类号: TP311.13

文献标志码: A

doi: 10.11805/TKYDA2018111

Visualization of urban taxi traffic flow based on graph structure

ZHANG Weiming¹, ZHANG Yong^{1,2}, LIU Hao², SUN Yanfeng¹

(1.Information Department, Beijing University of Technology, Beijing 100124, China;
2.Beijing Transportation Information Center, Beijing 100161, China)

Abstract: With the development of information technology, massive traffic data is exploited for traffic behavior analysis. In particular, taxi trajectory data plays an important role in traffic analysis. In order to better explore the traffic law of taxis in urban traffic operation, a graph based visual method is applied to the analysis of urban traffic flow in this paper. First, the road-level graph is converted into a region-level graph through clustering all road segments, then the node-line connection between regions presents the urban road network structure. Meanwhile, taking the regional traffic as the weight, the importance of the regions is analyzed by combining graph centrality. Finally, this approach is implemented and evaluated by using more than 100 million taxi trajectories data within the fourth ring of Beijing. The experiment results indicate that the proposed method can intuitively show the variation of taxi flow in different regions over time and reveal effectively the importance of different regions.

Keywords: taxi GPS; graphic structure; flow; visualization

随着我国城市化进程的加快, 城市人口急剧增加, 交通需求量也在快速增长。其中, 出租汽车交通在行人出行中承担着很重要的运输任务。与此同时, GPS 所提供的大量定位数据为出租车出行特征分析提供了有力的信息支撑。出租车轨迹数据不仅记录了真实的出租车位置信息, 还能反映城市交通运行状态。通过交互和可视化方法对这些轨迹数据进行分析, 不仅可以有效地帮助人们了解一个城市的交通运行规律状况, 同时为交通拥堵、行人出行分析提供便利。对出租车轨迹数据分析而言, 分析和理解这些轨迹数据的空间分布以及时间分布也是交通研究的主要内容。已有的关于出租车轨迹数据的研究工作大都针对交通运行状态分析, 如交通拥堵及扩散^[1]、交通状态预测^[2]等, 缺少对城市区域流量的宏观分析以及区域之间流量交互和区域运行状态分析, 对区域之间车流流向以及流量变化的可视化展现形式也较少。

为了解决以上问题, 本文提出了一种基于图结构化的路网表示方式, 将路网用点-线连接的图结构方式表示。同时结合图中心性概念, 对划分区域进行了重要性排序。这种可视化方法便于探索城市不同区域之间流量变化及流向, 反映宏观的城市交通规律。

收稿日期: 2018-06-01; 修回日期: 2018-07-31

基金项目: 北京市科技计划基金资助项目(Z171100000517003; Z171100000517004; Z161100001116072; Z171100004417023); 北京市教委基金资助项目(KM201510005025; KM201610005033)

作者简介: 张伟明(1992-), 男, 在读硕士研究生, 主要研究方向为轨迹数据可视分析、多媒体与智能软件技术。email:547875743@qq.com

1 相关工作

交通数据具有很强的空间相关性和局部性，对其进行空间划分是常用的可视化分析方法。目前对空间划分主要有屏幕像素划分、均匀网格划分、行政区域划分和对数据密度进行的多边形划分等^[3]。空间热力图(heat map)是像素划分的一种典型的可视化方法，通过像素点的颜色、深度的不同来表现数据的大小和变化趋势，它简单直接，应用也最为广泛。Andrienko 等^[4]将米兰划分成均匀网格区域，绘制了各区域的交通流方向分布和时间分布。Ferreira^[5]提出了一种支持大量 OD 数据和时空数据的分析系统，建立了一个查询模型，允许用户选择数据，并对纽约出租车数据进行详细分析，绘制了纽约市不同行政区的出租车上下车次数。Andrienko 等^[6]提出了一种基于数据空间分布的多边形划分方法，该方法对所有轨迹点进行密度聚类，通过聚类结果得到数据聚类位置并对其做了空间划分。Huang 等^[7]通过对出租车行驶轨迹进行路段拓扑关系构造，并对构造的道路拓扑结构进行 METIS 分区，通过图论中心性概念对区域进行重要性分析。Al-Dohuki 等^[8]提出一个语义交互可视化系统，通过文本转换将出租车轨迹转换为文本文档，将 GPS 轨迹点映射到一系列的街道名称，做空间文本分析。

在出租车轨迹可视化分析及应用方面，王祖超^[3]提出了一个交互式可视化系统分析城市交通拥堵状况，并根据 GPS 轨迹找到拥堵来源和扩散。Pan 等^[9]利用出租车旅客接送信息对土地利用进行分析和分类。Lu 等^[10]提出了一个可视化分析系统来帮助用户处理大规模的轨迹数据，比较不同的路线选择。Gao 等^[11]提出了一套针对出租车司机路线选择和决策的智能可视分析方法，用来分析出租车司机收入与驾驶行为的相关性。Liu^[12]等设计了一个 SmartAdP 可视化分析系统，通过对出租车轨迹数据分析来选取广告牌投放位置。除此之外，Dennis Thom^[13]等提出了一种行驶路线分析的可视化方法，通过选择起始位置和终止位置来分析可能选择的行驶路线。Sun 等^[14]提出了一种 Route-Zooming 可视化分析方法，将路段的时空信息嵌入到选择的路线上，并对比分析双向道路一周的交通状况。

在已有的研究中，已提出了很多对空间分区以及出租车轨迹数据可视化分析的方法及应用，但是对空间分区没有高效利用空间连接关系，同时对区域流量变化的宏观和中观相结合的可视化分析也比较少。

2 数据预处理及路网匹配

本文使用的地图是基于公开地图(Open Street Map, OSM)，而地图数据中存在一些不符合要求的道路，像人行道等道路是出租车无法行驶的，因此需要将人行道、未分类的相关道路删除。此外，将道路等级为高速、环路、一级、二级、三级的这 5 种道路数据保留，并根据路段的组成原则，对选定区域的道路结构进行结构重组。通过节点之间数据的连接关系构造北京市路网，形成完整的路网信息。

本文采用王祖超提出的数据清洗方法^[7]对出租车 GPS 数据进行清洗之后，对 GPS 数据进行经纬度坐标系转换，使 GPS 数据可以与 OSM 数据匹配，并可以在地图上显示。

由于 GPS 数据中不包含路段信息，采用最近距离匹配算法进行出租车 GPS 数据与路网的匹配，获得路段出租车信息。每匹配完成一个 GPS 采样点数据后，对 GPS 数据进行更新，随后标定匹配完的 GPS 点所在的路段编号信息。同时，为了便于车辆行驶轨迹处理，将每辆出租车的 GPS 数据按照采样时间顺序排序。

因为每个出租车 GPS 采样点的匹配涉及到与所有道路的最短距离的计算，为了加快计算速度，对路网进行网格划分。将北京市路网划分为 30×30 的网格，根据路网最大和最小的经纬度数值，计算出每个网格的经纬度范围，同时在数据库中标定好对应路段所属于的网格编号。这样，每次计算 GPS 点与道路的最短距离时，只需计算 GPS 与对应网格道路的最短距离即可。

3 基于图结构的城市交通流量可视化

为了探究城市交通流量宏观到区域的时空规律，以及区域流量的变化，本文采用从整体到局部的可视化设计思想，可视化分析方法主要由 4 个视图组成，如图 1 所示：(1)视图为图结构化的路网；(2)视图采用折线图表示区域一天的流量变化；(3)视图通过桑基图来表示某一段时间内区域之间的流量交换；(4)视图采用地图显示区域对应的实际道路，同时(4)视图包含 3 个小视图(a), (b),(c)：(a)采用南丁格尔玫瑰图对比区域 7 天流量的差异；(b)采用柱状图对比区域一天流量的流入与流出；

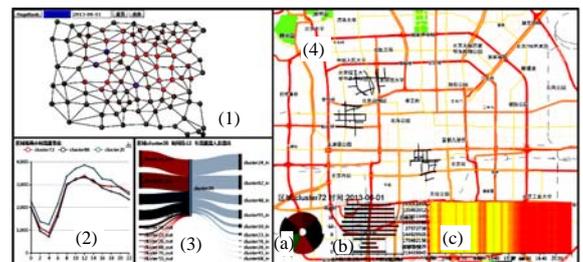


Fig.1 Diagram structured road network

图 1 图结构化路网

(c)采用热力图表示区域内路段每 10 min 的速度变化。

通过(1)视图将(4)视图中四环及以内的路段聚类结果进行图结构化展示，可以观察到区域之间的连接关系，通过节点颜色来宏观展示区域间流量差异、其余视图展示区域及区域之间在时空上的具体流量变化信息。

3.1 基于图结构的路网区域表示

为了表示区域之间的连接关系以及区域之间的相对位置，这里采用点—线连接的图结构方式表示路网结构。如图 2 所示，用 node 点表示聚类划分出的路网区域，点之间的连线 edge 表示划分区域之间的连接关系，最后将城市路网用无向图结构 $G<N,E>$ 表示。这样既反映了区域空间相对位置，又反映了区域连接关系，同时，也将道路级别路网转换成了区域级别。

节点的权重设定为区域出租车数量，边的权重设定为 1。同时使用图中心性概念，计算区域的 PageRank 数值。PageRank 是 Google 排名运算法则的一部分，是 Google 用来标识网页的等级/重要性的一种方法。本文通过出租车在区域之间的交互来计算区域的 PageRank 数值，进一步分析区域的重要性。如此计算得到的含义是：越高的 PageRank 值表明，这块区域是出租车采用的到其他地方的中心，比如市中心。

为了将路网图结构化，需将路网进行分区，分区的标准可以按路段的等级、流量、平均速度等，本文主要考虑区域之间的空间连接关系和流量交互规律，因此使用 K-means 聚类算法对路网经纬度信息进行聚类，将邻近路段聚类为同一区域，从而实现路网分区。

计算每个区域的中心经纬度信息并使用点代表区域，通过与屏幕坐标比例转换确定每个点之间的相对位置，点之间连接关系由实际路段的拓扑关系确定。当区域之间有多条连接路段时，只保留一条连接线，最终实现图结构化的路网，如图 1 中(1)视图所示。

采用图结构化的路网表示，既可以反映区域之间的连接关系，又可以直观观察城市整体出租车流量分布，同时可以直观地对比不同区域的流量差异。

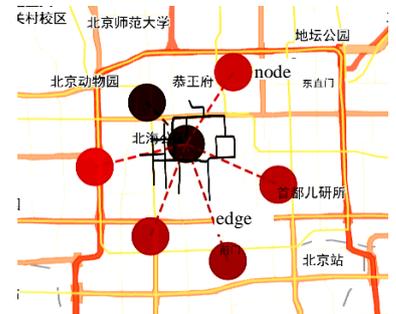


Fig.2 Visual analysis interface
图 2 可视化分析界面

3.2 基于桑基图的车流流向可视化

如图 1 中(3)视图所示，桑基图展示了某个区域在某个时间段内的出租车流量以及流向变化，左半部分表示该时段的流量主要来自哪些区域，右半部分表示流量流向哪些区域，延伸的分支宽度对应流量的大小。将流量变化折线图与桑基图关联，在折线图上选择想要分析的时间段。通过该视图可以知道某时段内与该区域进行流量交换的主要区域以及交换流量的大小。

3.3 区域流量及交通状态可视化

为了更详尽地分析单个区域在不同时间粒度(周，天，小时，分)的交通参数变化，本文采用南丁格尔玫瑰图、双向柱状图和热力图对其进行可视化。如图 3 所示，(a) 南丁格尔玫瑰图表示区域一周流量变化情况，花瓣面积越大，表示该天流量值越大。通过该视图可以发现区域流量一周变化规律以及是否在这一周内发生流量异常现象。(b) 双向柱状图展示了区域 1 天内 2 h 间隔的流入流量与流出流量对比，不仅可以看到流入与流出流量随时间的变化，还可以清晰地看出同一时间流入与流出的对比。(c) 热力图则用来展示对应区域内所有路段一整天的速度变化，设定时间间隔为 10 min，并且设定速度值低于 20 km/h 为红色，表示交通状态拥堵，速度值介于 20 km/h 到 40 km/h 之间为黄色，表示交通状态良好，大于 40 km/h 为绿色，表示交通状态通畅。通过图 3 的可视化展示，实现了在不同的时间粒度上对某个区域的流量及速度变化直观有效的展示。

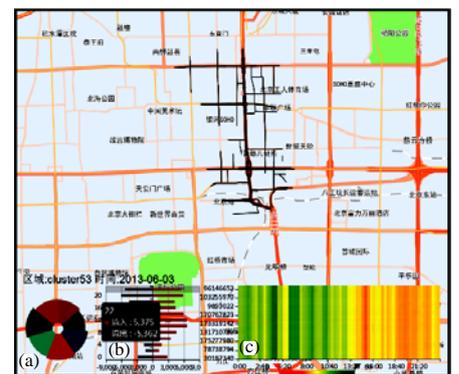


Fig.3 Regional traffic and traffic status visualization
图 3 区域流量及交通状态可视化

4 案例分析

本文使用的出租车数据由北京市交通信息中心提供，时间范围为 2013 年 6 月 1 日到 2013 年 6 月 7 日，每

天大约有 2 600 万多条数据。根据构造出来的路网结构，通过 *K*-means 聚类算法对路网经纬度信息进行聚类。*K* 值越大，划分的区域个数越多，区域包含的路段越少，反映的交通信息更加微观；*K* 值越小，划分的区域个数越少，区域包含的路段越多，反映的交通信息更加宏观。为了便于路网的可视化，本文将 *K* 值设为 100，将北京市四环以内路网划分为 100 个区域，将这 100 个区域采用图结构化方式表示，如图 4 所示，每一个点代表一个区域，点的颜色表示该区域一整天流量或 PageRank 大小。绿色表示数值小，红色表示数值大，同时颜色深浅表示数值从大到小。点上的数字表示区域的编号，与聚类结果对应，本文将路网聚类结果记为 cluster0 到 cluster99，即区域编号为 0~99。

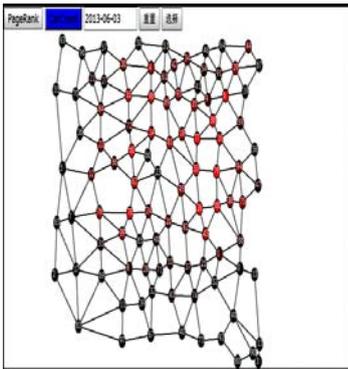


Fig.4 Diagram structured road network
图 4 图结构化路网

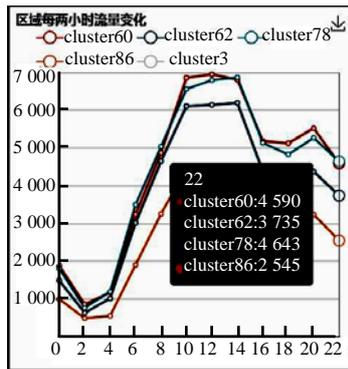


Fig.5 Area flow chart
图 5 区域流量折线图

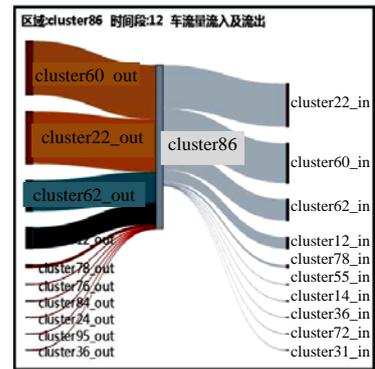


Fig.6 Area 86 flow interaction of Sankey view
图 6 区域 86 流量交互桑基图

4.1 区域流量可视化交互分析

以区域 60,62,78,86 这 4 个区域为例进行分析，在空间位置上，区域 60,62 和 78 为 3 个相邻区域，并且它们的路段以西二环环路为主。从图 5 的折线图可以发现这 3 个区域的流量变化相似，而区域 86 位于西二环与西三环之间，以一级路段和三级路段为主，区域的一天流量比前 3 个区域少，环路承担较多的出租车流量。

如图 6 所示，通过对区域 86 不同时间段的流量交互发现，流量交互主要以相邻区域 22,60 为主，与其他相邻区域的流量交互较少，而区域 22 主要路段为西三环环路，空间位置与区域 86 相连，可以推测出租车在西二环与西三环邻近区域之间，更多地选择通过区域 86 到达区域 22 和区域 60。

4.2 区域中心性分析

对区域 72,86 进行中心性分析，并选取区域 20 做对比，如图 7 所示，发现区域 72 和 86 的 PageRank 数值相似，而区域 20 的 PageRank 高于这 2 个区域，结合实际路网发现，区域 20 包括天安门广场、前门、大栅栏等，该区域是北京市中心，而区域 72 和 86 位于西二环和西三环之间，流量交换功能较弱于市中心，所以中心性低于区域 20。

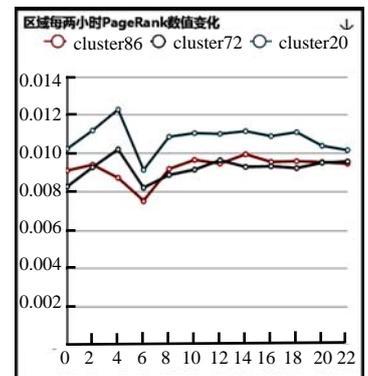


Fig.7 Rank line chart of area page
图 7 区域 PageRank 折线图

5 结论

本文提出了一种基于图结构的出租车交通流量分析方法，实现了路网分区以及可视化，并用于分析区域交通流量交互及变化规律。采用图结构表示路网，结合了 4 种可视化视图，使用户直观地观察到路网的区域划分和不同区域之间的流量交换，有效地揭示不同地区流量的变化。在将来的工作中，将进一步研究使用其他的分区算法，结合考虑区域兴趣点等因素，并对分区个数进行动态选定，对相邻区域的交通状态以及区域之间的流量交换做分析，探索区域之间变化影响因素，同时，分析不同区域内路段类别占比对交通流的影响，选择更有效的可视化分析方法。

参考文献：

[1] WANG Zuchao,LU Min,YUAN Xiaoru,et al. Visual traffic jam analysis based on trajectory data[J]. IEEE Transactions on Visualization & Computer Graphics, 2013,19(12):2159-2168.

-
- [2] ALVAREZ-Marquez A,AGUILERA I,GENTIL M A,et al. Traffic flow prediction for road transportation networks with limited traffic data[J]. IEEE Transactions on Intelligent Transportation Systems, 2015,16(2):653–662.
- [3] 王祖超,袁晓如. 轨迹数据可视分析研究[J]. 计算机辅助设计与图形学学报, 2015(1):9–25. (WANG Zuchao,YUAN Xiaoru. Visual analysis of trajectory data[J]. Journal of Computer-Aided Design & Computer Graphics, 2015(1):9–25.)
- [4] ANDRIENKO G,ANDRIENKO N. Spatiotemporal aggregation for visual analysis of movements[C]// Proceedings of IEEE Symposium on Visual Analytics Science and Technology. Los Alamitos:IEEE Computer Society Press, 2008:51–58.
- [5] FERREIRA N,POCO J,VO H,et al. Visual exploration of big spatio-temporal urban data: a study of new york city taxi trips[J]. IEEE Transactions on Visualization & Computer Graphics, 2013,19(12):2149–2158.
- [6] ANDRIENKO N,ANDRIENKO G. Spatial generalization and aggregation of massive movement data[J]. IEEE Transactions on Visualization and Computer Graphics, 2011,17(2):205–219.
- [7] HUANG Xiaoke,ZHAO Ye,YANG Jing,et al. TrajGraph:a graph-based visual analytics approach to studying urban network centralities using taxi trajectory data[J]. IEEE Transactions on Visualization & Computer Graphics, 2016,22(1):160–169.
- [8] AL-DOHUKI S,WU Y,KAMW F,et al. SemanticTraj:a new approach to interacting with massive taxi trajectories[J]. IEEE Transactions on Visualization & Computer Graphics, 2016,23(1):11–20.
- [9] PAN Gang,QI Guande,WU Z,et al. Land-use classification using taxi GPS traces[J]. IEEE Transactions on Intelligent Transportation Systems, 2013,14(1):113–123.
- [10] LU M,LAI C,YE T,et al. Visual analysis of multiple route choices based on general GPS trajectories[J]. IEEE Transactions on Big Data, 2017,3(2):234–247.
- [11] GAO Y,XU P,LU L,et al. Visualization of taxi drivers' income and mobility intelligence[M]. Berlin Heidelberg:Springer, 2012.
- [12] LIU D,WENG D,LI Y,et al. SmartAdP:visual analytics of large-scale taxi trajectories for selecting billboard locations[J]. IEEE Transactions on Visualization & Computer Graphics, 2017,23(1):1–10.
- [13] KRÜGER R,THOM D,WÖRNER M,et al. Trajectory Lenses—A set-based filtering and exploration technique for long-term trajectory data[J]. Computer Graphics Forum, 2013,32(3):451–460.
- [14] SUN G,LIANG R,QU H,et al. Embedding spatio-temporal information into maps by route-zooming[J]. IEEE Transactions Vis Comput Graph, 2017,23(5):1506–1519.